Metadata of the chapter that will be visualized in SpringerLink

Book Title	Artificial Intelligence in Education	
Series Title		
Chapter Title	Using Fair AI with Debiased Network Embeddings to Support Help Seeking in an Online Math Learning Platform	
Copyright Year	2021	
Copyright HolderName	Springer Nature Switzerland AG	
Author	Family Name	**Li**
	Particle	
	Given Name	**Chenglu**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	University of Florida
	Address	Gainesville, FL, USA
	Email	li.chenglu@ufl.edu
Corresponding Author	Family Name	**Xing**
	Particle	
	Given Name	**Wanli**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	University of Florida
	Address	Gainesville, FL, USA
	Email	wanli.xing@coe.ufl.edu
Author	Family Name	**Leite**
	Particle	
	Given Name	**Walter**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	University of Florida
	Address	Gainesville, FL, USA
	Email	walter.leite@coe.ufl.edu

Abstract

There has been a long-standing issue of sparse discussion forums participation in online learning, which can impede students' help seeking practices. Researchers have examined AI techniques such as link prediction with network analysis to connect help seekers with help providers. However, little is known whether these AI systems will treat students fairly. In this study, we aim to start a foundation work to build

	a recommender system that can (1) fairly suggest peers who are likely to answer a question and (2) predict the response quality of students.
Keywords (separated by '-')	Fair AI - Link prediction - Recommender system

Using Fair AI with Debiased Network Embeddings to Support Help Seeking in an Online Math Learning Platform

Chenglu Li, Wanli Xing(✉), and Walter Leite

University of Florida, Gainesville, FL, USA
li.chenglu@ufl.edu, {wanli.xing,walter.leite}@coe.ufl.edu

Abstract. There has been a long-standing issue of sparse discussion forums participation in online learning, which can impede students' help seeking practices. Researchers have examined AI techniques such as link prediction with network analysis to connect help seekers with help providers. However, little is known whether these AI systems will treat students fairly. In this study, we aim to start a foundation work to build a recommender system that can (1) fairly suggest peers who are likely to answer a question and (2) predict the response quality of students.

AQ1

Keywords: Fair AI · Link prediction · Recommender system

1 Introduction and Related Work

The discussion forum in online learning has been demonstrated to be an important learning tool given its collaborative nature that enhances learning through knowledge exchange [1,2]. However, there has been a long-standing issue of discussion forums that peer interactions are sparse [6]. The inactive use of discussion forums in online settings can impede students' help seeking practices [13]. Help seeking is an important skill in self-regulated learning and can positively affect students' learning outcomes [13,17,20]. To support students' help seeking in online discussion forums at a large scale, researchers have examined AI techniques such as link prediction with network analysis to connect help seekers with help providers [10,12,16]. Other than using network analysis such as structural similarity for link prediction, network embedding has recently shown to be a strong candidate [24]. Network embedding represents nodes in a graph with latent vectors such that neighboring nodes would have high similarity scores [19]. Studies have shown that network embedding can outperform prior link prediction algorithms [9,19,24].

While promising results on predictive accuracy have been presented in prior studies on automatically supporting help seeking in discussion forums, little is known whether these AI systems will treat students fairly. Algorithmically, studies have shown that AI can reflect humans' hidden values due to the existing

I. Roll et al. (Eds.): AIED 2021, LNAI 12749, pp. 1–6, 2021.
https://doi.org/10.1007/978-3-030-78270-2_44

bias in training datasets. For example, Caliskan et al. [5] found word embedding algorithms can perpetuated cultural stereotypes (e.g., females are highly correlated with family-oriented careers). Empirically, biases in AI have been found in domains such as education, hiring, and finance, where participants with specific demographics can be favored by predictive models [3,8,22]. In the case of link prediction, students might form communities of specific demographics. For example, white students dominantly interact with other white students because they come from the same school where minority students are scarce. Trained with such a dataset, models can reinforce the status quo and not give students opportunities to establish diversified connections that can be equally helpful. Therefore, to make AI in education sustainable, researchers need to purposefully address fairness issues [21,23]. In this study, we aim to start a foundation work to build a recommender system that can (1) fairly suggest peers who are likely to answer a question and (2) predict the response quality of students.

2 Methods

2.1 Research Context and Dataset

This study uses students' discussion forum, demographics, and log data on Algebra I from Algebra Nation (AN), an online math learning platform originated in Florida. The dataset consists of 17,794 post-reply pairs by 3,726 students with over 6 million logs in the academic year of 2018–2019. Post-reply pairs include contents of post and reply, poster IDs, and replier IDs. The log data captured students' interactions with AN (e.g., lecturing videos, reviewing videos, and discussion board).

2.2 Model Procedure

Link Prediction with Network Embeddings. Link prediction models are trained with network embeddings to predict if two students will be connected. For the network embeddings, we have examined Node2Vec [9] and DeBayes [4]. Node2Vec is inspired by the widely-applied algorithm Word2Vec [18]. In Node2Vec, nodes are analogous to words in Word2Vec, and the random walks algorithm is used to construct sequences of nodes. Latent vectors (embeddings) of nodes are then extracted from a neural network's hidden layer trained with the sequences. Node2Vec is selected because previous studies have achieved desired link prediction results with it. However, Node2Vec is fairness-unaware. To ensure the fairness of link prediction, we have also examined DeBayes modified based on Conditional Network Embeddings (CNE) [14] to learn fair representations. Conceptually, CNE solves for

$$P(G|X) = \frac{P(X|G)P(G)}{P(X)} \tag{1}$$

by finding an embedding X using Maximum Likelihood estimation, where G is the given network. Thus, the embedding will only need to capture information

that is NOT represented by the prior $P(G)$. DeBayes utilizes this property to get debiased embeddings by introducing a biased prior, where sensitive information related to protected groups is retained in the biased prior so that embeddings are not aware of such information.

Representation Bias and Equalized Odds. To evaluate fairness, we have examined representation bias (RB) [25] of network embeddings and equalized odds (EO) [11] in terms of **gender** and **races**. RB is the weighted average AUC scores of using embeddings to predict sensitive attributes (e.g., gender). Conceptually, embeddings are fair when RB is close to 0.5 since an AUC of 0.5 suggests a random classifier and we cannot infer students' sensitive information from embeddings. EO is defined as

$$P(\hat{Y} = 1 \, A = 0, Y = 1) = P(\hat{Y} = 1 \, A = 1, Y = 1) \tag{2}$$

$$P(\hat{Y} = 1 \, A = 0, Y = 0) = P(\hat{Y} = 1 \, A = 1, Y = 0) \tag{3}$$

, where $\hat{Y}$ is the predicted outcome of the model, Y is the binary outcome from the dataset (e.g., connected or not), and A is the comparison group (e.g., female vs. male). EO is satisfied when Eqs. 2 and 3 are met.

Response Quality Prediction. We have conducted a multiple linear regression analysis to understand what contributes to response quality. There were 27 predictors, which were repliers' standardized frequencies of interactions on Algebra Nation. Variance inflation factors (VIF) were calculated to avoid multicollinearity. Response quality of a reply is calculated based on its linguistic features (number of words and number of named entities), reputations (number of up-votes), readability (Flesch reading ease [15]), and coherence (cosine similarity between post and reply using BERT embeddings [7]). Log-transformation was applied to linguistic features and readability as we think the contribution of them decays as their values increase.

3 Results

Link Prediction. We evaluated models' predictive accuracy with AUC. The results show that Node2Vec has an AUC of 0.88 and DeBayes achieves that of 0.94. For the fairness evaluation (see Fig. 1), the representation bias of Node2Vec's embedding is 0.54 for gender and 0.53 for the race and that of DeBayes is 0.49 for gender and 0.5 for the race. In terms of equalized odds (EO), lower is fairer. Node2Vec has an EO of 0.037 for gender and 0.038 for race, while DeBayes has an EO of 0.002 for gender and 0.005 for race. The results suggested that DeBayes greatly outperformed Node2Vec in predictive accuracy and fairness.

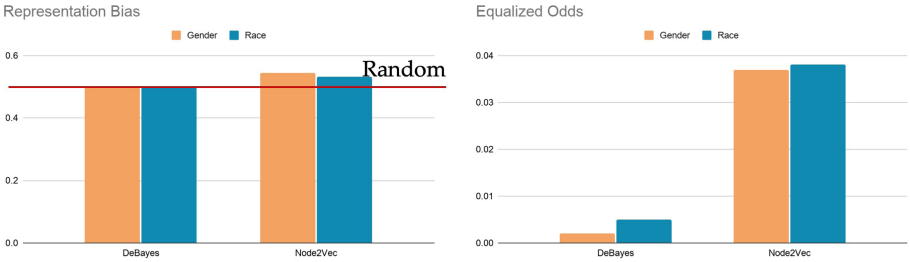

Fig. 1. Fairness evaluation of network embeddings and link prediction.

Table 1. Regression analysis results of the significant predictors

	Coef	P-value	Definition
Load discussions	.0528	<.000	Load the discussion forum page
Answer assessment	.0300	.002	Answer an assessment item
Finish assessment	−.1357	.016	Finish a whole assessment
Review incorrect assessment	−.0223	.006	Review the solution of an incorrect item
Create post	−.3183	<.000	Create a post in the discussion forum
Search discussions	.1714	.044	Search within the discussion forum
View document	.3191	.001	View learning resource files

Response Quality Prediction. The regression model demonstrates that 7 behaviors in discussion forums, video watching, and assessment taking are significant predictors of response quality. Table 1 illustrates the regression results and the significant predictors' definitions.

4 Conclusion

This paper has shown the possibility of conducting link prediction fairly while producing desirable accuracy. Although the fairness evaluation of Node2Vec does not suggest that the model is highly biased in our context, unlike DeBayes, Node2Vec is fairness-unaware and potential equity issues can arise without careful handling. Meanwhile, the regression analysis sheds light on the factors contributing to response quality. From a learning perspective, these significant predictors' effects on response quality are reasonable, indicating the validity of the computed response quality. In the future, we intend to triangulate the reliability and validity of the computed response quality with qualitative approaches.

Funding. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C160004 to the University of Florida and University of Florida AI Catalyst Grant. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

1. Almatrafi, O., Johri, A., Rangwala, H.: Needle in a haystack: identifying learner posts that require urgent response in mooc discussion forums. Comput. Educ. **118**, 1–9 (2018)
2. Biasutti, M.: A comparative analysis of forums and wikis as tools for online collaborative learning. Comput. Educ. **111**, 158–171 (2017)
3. Binns, R.: Fairness in machine learning: lessons from political philosophy. In: Conference on Fairness, Accountability and Transparency, pp. 149–159 (2018)
4. Buyl, M., De Bie, T.: Debayes: a bayesian method for debiasing network embeddings. In: International Conference on Machine Learning, pp. 1220–1229. PMLR (2020)
5. Caliskan, A., Bryson, J.J., Narayanan, A.: Semantics derived automatically from language corpora contain human-like biases. Science **356**(6334), 183–186 (2017)
6. Chiu, T.K., Hew, T.K.: Factors influencing peer learning and performance in mooc asynchronous online discussion forum. Australas. J. Educ. Technol. **34**(4) (2018)
7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186 (2019)
8. Giang, V.: The potential hidden bias in automated hiring systems. The Future of Work. Fast Company (2018)
9. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 855–864 (2016)
10. Hansen, P., et al.: Predicting the timing and quality of responses in online discussion forums. In: 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), pp. 1931–1940. IEEE (2019)
11. Hardt, M., Price, E., Srebro, N.: Equality of opportunity in supervised learning. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 3323–3331 (2016)
12. Howley, I., Tomar, G., Yang, D., Ferschke, O., Rosé, C.P.: Alleviating the negative effect of up and downvoting on help seeking in MOOC discussion forums. In: Conati, C., Heffernan, N., Mitrovic, A., Verdejo, M.F. (eds.) AIED 2015. LNCS (LNAI), vol. 9112, pp. 629–632. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19773-9_78
13. Howley, I., Tomar, G.S., Ferschke, O., Rosé, C.P.: Reputation systems impact on help seeking in mooc discussion forums. IEEE Trans. Learn. Technol. (2017)
14. Kang, B., Lijffijt, J., De Bie, T.: Conditional network embeddings. In: 7th International Conference on Learning Representations, ICLR 2019, p. 16 (2019). https://openreview.net/forum?id=ryepUj0qtX
15. Kincaid, J.P., Fishburne Jr., R.P., Rogers, R.L., Chissom, B.S.: Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Tech. rep, Naval Technical Training Command Millington TN Research Branch (1975)
16. Lan, A.S., Spencer, J.C., Chen, Z., Brinton, C.G., Chiang, M.: Personalized thread recommendation for MOOC discussion forums. In: Berlingerio, M., Bonchi, F., Gärtner, T., Hurley, N., Ifrim, G. (eds.) ECML PKDD 2018. LNCS (LNAI), vol. 11052, pp. 725–740. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-10928-8_43

17. Magnusson, J.L., Perry, R.P.: Academic help-seeking in the university setting: the effects of motivational set, attributional style, and help source characteristics. Res. High. Educ. **33**(2), 227–245 (1992)
18. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
19. Nelson, W., Zitnik, M., Wang, B., Leskovec, J., Goldenberg, A., Sharan, R.: To embed or not: network embedding as a paradigm in computational biology. Front. Genet. **10**, 381 (2019)
20. Newman, R.S.: How self-regulated learners cope with academic difficulty: the role of adaptive help seeking. Theory Pract. **41**(2), 132–138 (2002)
21. Pedro, F., Subosa, M., Rivas, A., Valverde, P.: Artificial intelligence in education: Challenges and opportunities for sustainable development (2019)
22. Riazy, S., Simbeck, K.: Predictive algorithms in learning analytics and their fairness. DELFI **2019** (2019)
23. Vincent-Lancrin, S., Van der Vlies, R.: Trustworthy artificial intelligence (ai) in education: Promises and challenges (2020)
24. Xu, Z., Ou, Z., Su, Q., Yu, J., Quan, X., Lin, Z.: Embedding dynamic attributed networks by modeling the evolution processes. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 6809–6819 (2020)
25. Zemel, R., Wu, Y., Swersky, K., Pitassi, T., Dwork, C.: Learning fair representations. In: International Conference on Machine Learning, pp. 325–333. PMLR (2013)

Author Queries

Chapter 44

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	